DOCUMENT RESUME

ED 393 870 TM 024 735

AUTHOR McTighe, Jay; Ferrara, Steven
TITLE Assessing Learning in the Classroom. A Report from Professional Standards and Practice.
INSTITUTION National Education Association, Washington, D.C.
PUB DATE Nov 94
NOTE 35p.
AVAILABLE FROM National Education Association, Professional Standards and Practice, 1201 16th Street, N.W., Washington, DC 20036-3290.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Constructed Response; *Educational Assessment; Educational Improvement; Educational Practices; Elementary Secondary Education; *Evaluation Methods; Formative Evaluation; Higher Education; Learning; Objective Tests; *Performance Based Assessment; Reliability; Selection; *Standards; *Student Evaluation; Summative Evaluation; *Test Construction; Validity
IDENTIFIERS *National Education Association

ABSTRACT

A variety of methods are examined that teachers from preschool to graduate school levels can use in assessing their students; the common principles underlying classroom assessment are explored. The first principle is that the primary purpose of classroom assessment is to inform teaching and improve learning. A second principle is that multiple sources of information are necessary when assessing learning in the classroom. A third principle of classroom assessment concerns validity, reliability, and fairness. Once these principles are accepted, the selection of particular assessment methods should be based on desired learning outcomes, the purpose of the assessment, and audience for which it is intended. Assessment approaches that might be used include selected response items of the sort presented in multiple-choice, true-false, and matching tests and for performance-based approaches that include constructed responses, product assessment, performances, and process-focused assessment. In addition to making choices about classroom assessment methods, teachers should consider options for evaluating student work and for communicating assessment results. Various scales and reporting processes are discussed. An appendix contains a glossary of classroom assessment terms. (Contains 6 figures and 22 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 393 870

A REPORT FROM PROFESSIONAL STANDARDS AND PRACTICE

Assessing Learning In the Classroom

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY

G. FELTON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

nea
NATIONAL EDUCATION ASSOCIATION

PROFESSIONAL STANDARDS AND PRACTICE
1201 16th Street, NW
Washington, DC 20036-3290

(202) 822-7350
FAX (202) 822-7482

BEST COPY AVAILABLE

TM024735

Contents

Introduction 3

Assessing Learning in the Classroom 4

- Terminology 4
- Large Scale versus Classroom Assessment 6
- Principles of Effective Classroom Assessment 7
- Key Questions in Planning Classroom Assessments 10
- The Framework of Assessment Approaches and Methods . 12
- Evaluation and Communication Methods 21
- Conclusion 28

References 29

Appendix A: Glossary of Classroom Assessment Terms ... 31

INTRODUCTION

Delegates to the 1993 NEA Representative Assembly adopted a new business item that called upon the NEA to "implement a process for identifying the best and fairest assessment practices and disseminate information on these practices to NEA affiliates and members." They additionally called for a 1994 progress report to the NEA Board of Directors "on assessment practices and the role of high-stakes testing."

In response to this new business item and in consultation with the 1993-94 NEA Standing Committee on Professional Standards and Practice, the NEA Professional Standards and Practice unit commissioned this paper on classroom-based assessment of learning from Maryland testing experts Jay McTighe and Steven Ferrara. The authors hold the view that "the primary purpose of classroom assessment is to inform teaching and improve learning, not to sort and select students or to justify a grade." They discuss principles of effective classroom assessment, illustrate a variety of assessment approaches and methods, and provide a framework for planning. Their readable, practical approach to an evolving and sometimes daunting subject is intended to be of use to teachers at all levels, preschool through graduate studies.

Jay McTighe directs the Maryland Assessment Consortium, a collaboration among Maryland's school systems to develop and share performance assessments with teachers throughout the state. He is well known for his work with "thinking skills," having coordinated statewide efforts to develop instructional strategies, curriculum models, and assessment procedures for improving the quality of student thinking.

Steven Ferrara, Ph.D., is Maryland's State Director of Student Assessment. A former high school teacher, he also has taught at the preschool, middle school, and graduate levels. He has spent the last 10 years conducting research, writing, teaching, and speaking on the effects of large-scale and classroom assessment on learning.

The authors wish to thank the following individuals for their comments on and critique of various drafts of this document: Gail Goldberg, Joan Heiss, Mary Kimmins, Hillary Michaels, Ruth Mitchell, and Barbara Reeves.

Assessing Learning in the Classroom

On-going assessment of student learning in the classroom is an essential aspect of effective teaching. Teachers can use a variety of assessment methods to diagnose students' strengths and needs, plan and adjust instruction, and provide feedback to students and parents regarding progress and achievement. In this report, we take the position that the primary purpose of classroom assessment is to inform teaching and improve learning, not to sort and select students or to justify a grade.

In this report, we take the position that the primary purpose of classroom assessment is to inform teaching and improve learning, not to sort and select students or to justify a grade.

The report is intended for teachers from the preschool to graduate school levels to use in examining a variety of methods for effectively and fairly assessing their students. While the choice of particular assessment methods will vary according to the purpose of the assessment, the content of the curriculum, and the age levels of students, a set of common principles underlies effective classroom assessment. We discuss these principles, consider the strengths and limitations of a variety of assessment approaches and methods, and present a series of vignettes to illustrate classroom assessment in action. Finally, we offer a set of guiding questions and a framework for planning classroom assessments to improve teaching and learning.

Terminology

Teachers frequently begin new units of study by introducing or reviewing key vocabulary with the recognition that an understanding of certain basic concepts will enhance subsequent learning of important principles and procedures in the unit. Likewise, we'll begin with a review of basic terminology commonly associated with classroom assessment.

Assessment refers to *"any systematic basis for making inferences about characteristics of people, usually based on various sources of evidence; the global process of synthesizing information about individuals in order to understand and describe them better"* (Brown, 1983). While considering this textbook definition, it is interesting to note that the term *assessment* is derived from the Latin root *assidere* meaning "to sit beside." While this original meaning

may seem at odds with present images of large-scale, standardized testing—number 2 pencils, "bubble" sheets, rigid time limits, silent work, etc.—it conforms more closely with the array of assessment methods routinely used by teachers for assessing their students. *Assidev* suggests that, in addition to tests and projects, classroom assessments include informal methods of "sitting beside," observing, and conversing with students as a means *of understanding and describing what they know and can do.*

The terms, *assessment, testing,* and *evaluation,* are frequently used interchangeably, although they have distinct meanings. *Assessment* is a broad term referring to the process of gathering and synthesizing information to help us understand and describe. *Testing* is one type of assessment. Tests generally utilize a paper-and-pencil format, occur within established time limits, restrict access to resources (e.g., reference materials), and yield a limited range of acceptable responses. *Evaluation* involves making a judgment regarding quality, value, or worth, based upon criteria. Teacher questioning, reviews of student work folders, and paper-and-pencil tests are widely used assessment methods for gathering information about student learning. Scoring a student essay or assigning a report card grade are examples of evaluation.

The purposes and audiences for assessment information influence what is assessed, how it is assessed, and how the results are communicated and used.

Another pair of widely used terms, *formative* and *summative* assessment, pertain to the purpose and timing of classroom assessments. *Summative assessment* refers to any culminating assessment that provides a summary report on the degree of knowledge or proficiency attained at the conclusion of a unit, course, or program of study. A final exam, senior exhibition, or dissertation defense are examples of summative assessments. *Formative assessments* refer to ongoing, diagnostic assessments that provide information to help teachers adjust instruction and improve student performance. For instance, prior to the start of a unit on the Civil War, a teacher might ask students to make a "web" or an outline to show what they already know about this period of history as a means of obtaining information about students' prior knowledge. She might also interview several students at random to check their Civil War awareness. Formative assessment can also be used during instruction to check on student understandings and misconceptions. Teachers often use brief written and oral quizzes and classroom discussions to determine if students have learned course material and can apply instructed skills. Such activities provide teachers with valuable information that allows them to adjust instruction to achieve more effective learning.

The term *alternative* assessment appears widely in the current education

literature (Herman, Aschbacher, & Winters, 1992). While there is not a universally agreed-upon definition for the term *alternative*, it generally refers to assessments that differ from the multiple-choice, timed, "one-shot" approaches that characterize most standardized and some classroom assessments. We do not recommend the use of this term because it is imprecise and open to various interpretations.

We will introduce additional assessment terminology in the following sections. For definitions of related terms important in classroom assessment, see the glossary (Appendix A).

Large Scale versus Classroom Assessment

Different types of assessments exist to address different information needs. The purposes and audiences for assessment information influence what is assessed, how it is assessed, and how the results are communicated and used. Large-scale standardized tests, such as the Iowa Tests of Basic Skills (ITBS), California Achievement Tests (CAT), and the Stanford Achievement Test, are used primarily to satisfy the demands for educational accountability. The results of assessments such as these are reported to legislatures, boards of education, school administrators, parents, and the general public. Standardized tests are generally *norm referenced* to allow for easy interpretation. They are designed to determine how well students have learned particular concepts and skills as compared to other students in a norming group. The results of norm-referenced assessments may be conveniently displayed so that observers can readily distinguish achievement above or below the norm. Not all large-scale, standardized tests are norm-referenced. Some, such as the College Board's Advanced Placement Examinations and certain state-level competency tests, are *criterion referenced*, whereby student performance is evaluated and reported compared to pre-established standards. Likewise, not all standardized tests are multiple-choice in nature. Several states currently use standardized performance assessments, featuring open-ended tasks, for "high stakes" accountability purposes.

Because they are intended to provide accountability information, "one-shot," large-scale standardized tests typically do not provide sufficiently detailed or timely information regarding student achievement of specific curriculum goals.

Standardized tests are considered "high stakes" tests if their results are used for consequential decisions such as promotion, graduation, admission, certification, evaluation, or where rewards and sanctions are involved. For example, a district-wide minimum competency exam would be "high stakes" for students if passing the exam is a requirement for a high school

diploma. For an extended discussion of standardized testing, see *The Role of High-Stakes Testing in School Reform* (Smith, 1993). Because they are intended to provide accountability information, "one-shot," large-scale standardized tests typically do not provide sufficiently detailed or timely information regarding student achievement of specific curriculum goals.

At the classroom level, teachers have different assessment needs—diagnosing student strengths and weaknesses, informing students and parents about progress, planning and adjusting instruction, and motivating students to focus on valued knowledge and skills. With these purposes in mind, classroom assessments may be tailored directly to the curriculum and to the information needs of individual teachers, students, and parents. Unlike "one-shot," standardized tests, classroom assessments designed to promote learning are more likely to occur over time, include an array of methods, focus on elements of quality, offer a more personalized picture of student achievement, and provide timely and specific feedback.

Principles of Effective Classroom Assessment

A wide variety of methods is available to teachers for assessing student learning (Airasian, 1991; Cross & Angelo, 1988; Ferrara & McTighe, 1992). Regardless of the particular methods employed, effective classroom assessment is guided by three fundamental principles. Classroom assessment should: 1) promote learning; 2) use multiple sources of information; and 3) provide fair, valid, and reliable information.

Rather than waiting until the end of a unit of study or course to assess students, effective teachers employ formative assessments at the beginning of instruction to determine students' prior knowledge.

The first principle is based upon the premise that the primary purpose of classroom assessment is to **inform teaching and improve learning** (Mitchell and Neill, 1992). This premise suggests that classroom assessment be viewed as an on-going process instead of a single event at the conclusion of instruction. Rather than waiting until the end of a unit of study or course to assess students, effective teachers employ formative assessments at the beginning of instruction to determine students' prior knowledge. They assess regularly throughout the unit or course of study to obtain information to help them adjust their teaching based on the learning needs of students. They recognize that assessment results can inform them about the effectiveness of their teaching as well as the degree of student learning.

When using performance-based assessments, teachers can make their evaluative criteria explicit in advance to serve as a focus for both

instruction and evaluation. Effective teachers help their students understand that criteria describe the desired elements of quality. They provide regular feedback to students based on the identified criteria and allow students to revise their work based upon this feedback. They also involve students in peer- and self-evaluation using the criteria in order to engage students more actively in improving their performance.

Assessment for learning recognizes the mutually supportive relationship between instruction and assessment. Like a Möbius strip where one side appears to seamlessly blend into the other, classroom assessment should reflect and promote good instruction. For example, teachers following a process approach to teaching writing would allow their students to develop drafts, receive feedback, and make revisions as part of the assessment. Likewise, if we teach science through a hands-on, experimental approach, our assessment should include hands-on investigations.

Applying the principle of multiple sources is especially important when the assessment information is used as a basis for important summative decisions, such as assigning report card grades or determining promotion.

A second principle of sound classroom assessment calls for a synthesis of information from several sources. The importance of using **multiple sources of information** when assessing learning in the classroom may be illustrated through a photographic analogy. A single assessment, such as a written test, is like a snapshot in that it provides a picture of student learning. While a snapshot is informative, it is generally incomplete since it portrays an individual at a single moment in time within a particular context. It is inappropriate to use a one-time snapshot of student performance as the sole basis for drawing conclusions about how well a student has achieved desired learning outcomes. The classroom context offers a distinct advantage over large-scale assessments in that it allows teachers to take frequent samplings of student learning using an array of methods. To continue the photographic analogy, classroom assessment enables us to construct a "photo album" containing a variety of pictures taken at different times with different lenses, backgrounds, and compositions. The photo album reveals a richer and more complete picture of each student than any single snapshot can provide. Applying the principle of multiple sources is especially important when the assessment information is used as a basis for important summative decisions, such as assigning report card grades or determining promotion.

A third principle of classroom assessment concerns **validity**, **reliability**, and **fairness**. *Validity* has to do with whether an assessment measures what it was intended to measure. For example, if a media specialist seeks to

assess the capabilities of her students to conduct research using primary and secondary sources, she should observe their use of these sources directly as they work on their research projects. For this learning outcome, a paper-and-pencil test of student knowledge of library references would be an indirect and less valid assessment since it does not reveal the ability to actually use the references purposefully.

Reliability refers to the dependability and consistency of assessment results. If the same assessment yielded markedly different results with the same students (without intervening variables such as extra instruction or practice time), one would question its reliability. Performance assessments present an additional challenge since they call for judgment-based evaluation of student products and performances. A truly reliable evaluation would result in equivalent ratings by the same rater on different occasions. For instance, an observation checklist can be used reliably as long as teachers are careful to ensure that their ratings would not differ substantially from occasion to occasion (e.g., Monday morning versus Friday afternoon). When teachers are involved in school- or district-level evaluations based on a set of criteria used throughout the school or district, inter-rater reliability must also be considered. In this case, scores on a writing assessment would be considered reliable if different raters assign similar scores to the same essays.

Fairness *in classroom assessment refers to giving all students an equal chance to show what they know and can do. Fairness is compromised when teachers assess something that hasn't been taught or use assessment methods that are incongruent with instruction.*

Fairness in classroom assessment refers to giving all students an equal chance to show what they know and can do. Fairness is compromised when teachers assess something that hasn't been taught or use assessment methods that are incongruent with instruction (e.g., asking for recall of facts when the emphasis has been on reasoning and problem solving). The fairness of teacher judgments is also challenged by the "halo" and "pitchfork" effect, where expectations based on a student's past attitude, behavior, or previous performance influence the evaluation of his or her current performance.

Subtle, unintended racial, ethnic, religious, or gender biases also present roadblocks to the fair assessment of students. Such biases may negatively influence students' attitudes toward, and performances on, classroom assessments. For example, the junior high mathematics teacher who routinely uses sports statistics as a main source for problem-solving tasks could "turn off" those students who are not sports fans. Likewise,

insensitivity to diverse religious beliefs (e.g., choosing reading passages involving only Christian holidays), gender/racial images (e.g., depicting all doctors as white males), or socio-economic status (e.g., assuming that all kids have access to a telephone or home computer) may result in unfair evaluation of individuals or groups. Teachers must be on guard so that biases do not influence their evaluations of a student's performance.

Given these three general principles, we now consider a set of fundamental questions related to planning classroom assessments.

Key Questions in Planning Classroom Assessments

Just as teachers have numerous instructional techniques and strategies from which to choose, a variety of methods are available for assessing learning. The selection of particular assessment methods should be determined in response to several key questions. These questions are incorporated in the chart, "Classroom Assessment Planning: Key Questions" (Figure 1).

The first question concerns **learning outcomes**, or the intended results of our teaching: *What do we want students to understand and be able to do?* Learning outcomes typically fall into three categories: 1) *declarative knowledge*—what we want students to understand (facts, concepts, principles, generalizations); 2) *procedural knowledge*—what we want students to be able to do (skills, processes, strategies); and 3) *attitudes, values, or habits of mind*—how we would like students to be disposed to act (e.g., appreciate the arts, treat people with respect, avoid impulsive behavior). The choice of specific assessment methods should be determined in large part by the nature of the learning outcomes being assessed (Marzano, Pickering, & McTighe, 1993). For example, if we want students to demonstrate the capacity to write an effective persuasive essay, then our assessment should involve gathering samples of persuasive writing and evaluating them against specified criteria. In this case, a multiple-choice test would be ill-suited to measure the intended outcome. Likewise, if we wish to develop students' ability to work cooperatively on a research project, then we would assess group processes and products as well as individual performance.

In addition to considering outcomes, we need to raise questions related to the **purpose(s)** and **audience(s)** for classroom assessments: *Why are we assessing? How will the assessment information be used? For whom are the assessment results intended?* The purpose(s) and audience(s) for assessments influence not only the methods selected, but also the ways in which the results of classroom assessments are communicated. For example, if we wish to provide parents of a primary-grade student with an interim report

Figure 1
Classroom Assessment Planning: Key Questions

LEARNING OUTCOMES

What do we want students to understand and be able to do?

- ■ _____
- ■ _____
- ■ _____
- ■ _____

PURPOSE(S) FOR ASSESSMENT

Why are we assessing and how will the assessment information be used?

- ❑ to diagnose student strengths and needs
- ❑ to provide feedback on student learning
- ❑ to provide a basis for instructional placement
- ❑ to inform and guide instruction
- ❑ to communicate learning expectations
- ❑ to motivate; to focus student attention and effort
- ❑ to provide practice applying knowledge and skills
- ❑ to provide a basis for student evaluation (e.g., grading)
- ❑ to obtain data for site-based management
- ❑ to gauge program effectiveness

AUDIENCE(S) FOR ASSESSMENT

For whom are the assessment results intended?

- ❑ teacher / instructor
- ❑ students
- ❑ parents
- ❑ grade-level / department team
- ❑ other faculty
- ❑ school administrators
- ❑ curriculum supervisors
- ❑ business community
- ❑ higher education
- ❑ general public
- ❑ other: _____

of progress in language arts, we might arrange a conference to describe the child's reading skills in terms of a developmental profile and review a work folder containing samples of her writing.

The Framework of Assessment Approaches and Methods

Given identified outcomes, purposes, and audiences, *how might we assess student learning in our classrooms?* The "Framework of Assessment Approaches and Methods" (Figure 2) offers a systematic guide to the purposeful selection of assessment methods.

Each of the five columns in the Framework identifies an assessment approach and contains examples of specific assessment methods corresponding to that approach. In this section, we'll describe each of the five general approaches, examine the strengths and limitations of each, and provide classroom vignettes of teachers using various assessment methods.

Selected-Response Items

Selected-response formats—multiple choice, true-false, and matching—are widely known and used in educational testing, especially at the secondary and post-secondary levels (Haertel, Ferrara, Korpi, & Prescott, 1984; Stiggins & Conklin, 1992). Multiple-choice items are the most common type of selected responses, appearing on most commercially produced tests, as well as on many classroom assessments. The selected-response format presents students with a question, problem, or statement followed by a set of alternative responses. Students make a selection from among the given alternatives rather than generate their own response. While most selected-response items have a single correct, or best response, it is possible to create "enhanced choice" items that have more than one acceptable answer contained among the alternatives.

Selected-response formats allow teachers to efficiently and objectively assess students' knowledge of factual information, concepts and principles, and the application of basic skills. Because a large number of selected-response items may be included on a test, they enable a teacher to sample a broad range of knowledge and skills in a limited amount of time. Since selected-response items contain correct or acceptable responses, they are easily and objectively scored as correct or incorrect using an answer key. Machine-scorable answer sheets and hand-scoring templates simplify the scoring process, allowing teachers to quickly obtain results for timely feedback to students.

Despite their advantages, selected-response items have limitations. Instead of assessing the application of knowledge and higher-order skills in meaningful, "real world" situations, they tend to assess knowledge and

Figure 2

Framework of Assessment Approaches and Methods

How might we assess student learning in the classroom?

Selected-Response Items

- ❑ multiple choice
- ❑ matching
- ❑ true-false

PERFORMANCE-BASED ASSESSMENTS

Constructed Responses

- ❑ fill in the blank
  - word(s)
  - phrase(s)
- ❑ short answer
  - sentence(s)
  - paragraphs
- ❑ label a diagram
- ❑ "show your work"
- ❑ visual representation
  - web
  - concept map
  - flow chart
  - graph/table
  - matrix
  - illustration

Products

- ❑ essay
- ❑ research paper
- ❑ log/journal
- ❑ lab report
- ❑ story/play
- ❑ poem
- ❑ portfolio
- ❑ art exhibit
- ❑ science project
- ❑ model
- ❑ video/audiotape
- ❑ spreadsheet

Performances

- ❑ oral presentation
- ❑ dance/movement
- ❑ science lab demonstration
- ❑ athletic competition
- ❑ dramatic reading
- ❑ enactment
- ❑ debate
- ❑ musical recital

Process-Focused

- ❑ oral questioning
- ❑ observation ("kid watching")
- ❑ interview
- ❑ conference
- ❑ process description
- ❑ "think aloud"
- ❑ learning log

skills in isolated, decontextualized ways. Selected-response items cannot adequately measure certain learning outcomes, such as critical thinking, creativity, oral communication, and social skills. While real-world issues and problems rarely have single correct answers, the widespread use of selected-response assessments may communicate to students unintended messages about the nature of knowledge and learning; that is, recognizing the "right answer" is the primary goal of education. Critics also express concern that multiple-choice tests lead to "multiple-choice" teaching; that is, a focus on acquisition of facts rather than an emphasis on understanding and the thoughtful application of knowledge (Mitchell, 1992; Perrone, 1991; Wiggins, 1991). With a recognition of their advantages and limitations, teachers may appropriately incorporate selected-response items as part of a balanced menu of assessment approaches.

While real-world issues and problems rarely have single correct answers, the widespread use of selected-response assessments may communicate to students unintended messages about the nature of knowledge and learning; that is, recognizing the "right answer" is the primary goal of education.

Classroom Examples

Prior to introducing photosynthesis, a high school biology teacher gives her students a multiple-choice pretest to assess their knowledge of basic concepts related to the topic. Using the school's scanning machine to score the pretest, the teacher quickly determines the prior knowledge of students in her classes and plans her instruction accordingly.

A college economics professor includes a multiple-choice section on her mid-term and final examinations to assess students' knowledge of key concepts and principles presented in the text. The results are used in conjunction with other factors (essay questions, research paper, class participation) to derive the course grade.

At the beginning of an AIDS awareness unit, a middle school health teacher gives his students a true-false quiz to determine their knowledge of basic facts about the disease. Using a scoring key, the teacher quickly realizes that the majority of students hold several common misconceptions regarding the spread of AIDS. This information helps the teacher to address these misconceptions during the unit.

Upper elementary students take a practice quiz in which they match states with their capital cities. They then exchange papers with a partner and check each other's responses by consulting a United States map. Incorrect matches are identified for later review with a "study buddy."

The development of fair and valid selected-response tests is a challenging and time-consuming process. While a complete treatment of this topic

is beyond the scope of this publication, several excellent resources are available to assist teachers in designing and using selected-response format assessments for their classrooms. For more information, see Carlson, 1985; Nitko, 1983; Osterlind, 1989.

Performance-Based Assessment

The approaches in the second through fifth sections of the Framework may be categorized under the broader heading of performance-based assessment. By *performance-based assessment*, we are referring to assessment activities that directly assess students' understanding and proficiency. These assessments allow students to construct a response, create a product, or perform a demonstration to show what they understand and can do. Since they call for students to apply knowledge and skills rather than simply to recall and recognize, performance-based assessments are more likely to reveal student understanding. They are well suited to assessing application of content-specific knowledge, integration of knowledge across subject areas, and lifelong learning competencies such as effective decision making, communication, and cooperation (Shepard, 1989).

> *The term* authentic *is used to convey the idea that performance assessments should engage students in applying knowledge and skills in ways that they are used in the "real world."*

The current interest in performance-based methods has popularized additional assessment terms, such as *authentic*, *rubric*, *anchors*, and *standards*. The term *authentic*, popularized by Grant Wiggins (Wiggins, 1989), is used to convey the idea that performance assessments should engage students in applying knowledge and skills in ways that they are used in the "real world." According to Wiggins, authentic assessments should also reflect good instructional practice in ways that make "teaching to the test" legitimate and worthwhile.

Since performance-based assessments generally do not yield a *single* correct answer or solution method but allow for a wide range of responses, evaluations of student products or performances are based upon judgments. The evaluative judgments are guided by criteria that define the desired elements of quality. One widely used scoring tool is a *rubric*, a generic scoring tool used to evaluate the quality of products and performances in a given outcome area. Rubrics consist of a fixed measurement scale (e.g., 4-point) and a list of criteria that describe the characteristics for each score point. Rubrics are frequently accompanied by examples of student products or performances to illustrate each of the points on the scale. These examples are called *anchors*.

The term *standards* is frequently used in conjunction with performance-based assessments. There are three distinct ways in which the term is being used in the current educational discussions: 1) *content standards*, which specify *what* students should know and be able to do; 2) *performance standards*, which set expectations about *how well* students should perform; and 3) *opportunity-to-learn* standards, having to do with the necessary *resources and conditions* for effective teaching and learning. Performance-based assessments call for decisions about content standards as well as expected standards for performance (Diez, 1993).

Given these general considerations, we'll now take a closer look at specific approaches and methods of performance-based assessments.

Constructed-Response Formats

Unlike selected-response items that call for a selection from given alternatives, constructed-response assessment tasks ask students to generate brief responses to open-ended questions, problems, or prompts. Short written answers and visual representations (e.g., concept map, flow chart, graph) are examples of widely used constructed-response assessment methods. While constructed-response tasks may seek a correct or acceptable response (e.g., fill-in-the-blank), they are more likely to yield a range of responses. Thus, the evaluation of student responses requires judgment, guided by criteria. This approach may be used for assessing declarative knowledge and procedural proficiency. In addition, constructed responses can provide insight into understanding and reasoning when students are requested to show their work and explain or defend their answers in writing.

Constructed responses offer several advantages as classroom assessments. They require less time to administer than other types of performance-based assessments. Since they elicit short responses, a variety of constructed-response tasks may be used to assess multiple outcomes. Evaluation of student responses is straightforward, guided by criteria or model responses.

Constructed responses are limited in their ability to adequately assess attitudes, values, or habits of mind. In addition, as with any open-ended assessment, judgment-based evaluation takes time, and introduces potential problems of scoring reliability and fairness. For summative assessments, constructed-response tasks should not be regularly reused in order to avoid memorized responses to known questions and tasks.

Classroom Examples

> *A middle school science teacher involves her students in an investigation of the absorbency rates of different brands of paper towels. Following the investigation, they record results of their data collection on a chart and*

state their conclusions in writing. Students are evaluated on the effectiveness of their charts in communicating results and on the accuracy of their written conclusions.

As part of a culminating assessment for a social studies unit, fourth graders are given a blank map of their state. They are asked to fill in the names of their state's capital and the three other largest cities, and label the two major rivers.

As part of a course review, a high school psychology teacher provides his students with a comparison matrix containing different psychological theories on one axis and various personality disorders on the other. They are asked to fill in the cells of the matrix to explain the personality disorders according to each theory. The teacher then reviews the matrices and discusses them with the class to clear up misunderstandings.

A sixth grade mathematics teacher asks her students to show their work as they attempt to solve multistep word problems. In addition to examining their solutions, she looks at their work for evidence of appropriate use of algorithms and problem-solving strategies. She provides feedback through brief written comments.

Product Assessments

Student products provide tangible indicators of the application of knowledge and skills. Many educators believe that product assessment is especially "authentic" because it closely resembles the real work outside of school. Teachers may evaluate **written** products (e.g., essays, research papers, laboratory reports), **visual** products (e.g., 2- and 3-dimensional models, displays, videotapes), **aural** products (e.g., an audiotape of an oral presentation), and other types of products to determine degrees of proficiency or levels of quality.

Many teachers recognize that evaluation criteria also serve an instructional purpose: providing students with a clear focus on elements of quality to guide their work.

Product assessment calls for the selection or development of criteria for evaluation. The criteria are incorporated into a generic scoring rubric, task-specific guide, rating scale, or checklist. Many teachers recognize that evaluation criteria also serve an instructional purpose: providing students with a clear focus on elements of quality to guide their work. When the criteria are made public, students may be involved in using them in peer- and self-evaluation of products.

One application of product assessment occurs when representative samples of student work are systematically collected over time in portfolios. Portfolios allow teachers, students, parents, and others to observe development and growth in learning. Portfolio assessment has been widely

used over the years in the visual arts, architecture, and certain technical areas.

Currently, educators are witnessing a growing use of portfolios to document learning in other subject areas, especially the language arts. For additional information on the classroom use of portfolios, see *Student Portfolios* (National Education Association, 1993).

The use of products and portfolios is appealing because of their instructional relevance. When students are given opportunities to produce authentic products, they often become more engaged in, and committed to, their learning. Unlike standardized assessments that strive for uniformity, product assessment often presents students with opportunities to express their individuality. Product assessment highlights what students **can do**, while revealing what they need to learn or improve. The criteria used to evaluate products, when made public, make the elements of quality known to students and serve as a guide for student peer- and self-evaluation. Previously developed products can serve an instructional purpose when they are presented as models of excellence for students (Wiggins, 1992).

Despite their benefits, product assessments have their drawbacks. Criteria for judging the products must be identified, and product evaluation can be a time-consuming process. In addition, teachers must be careful when evaluating student products that their judgments aren't unduly influenced by extraneous variables, such as neatness or spelling. Practicality must also be considered. The time required to develop quality products may compete with other instructional priorities. Product assessments require resources, including funds for materials and space for display and storage.

Classroom Examples

Students develop a computer program for an advanced high school computer class. Their teacher evaluates students' programming knowledge and skills by examining the program's written code for accuracy and efficiency. In addition, students must run the program to demonstrate that it performs the specified functions. Unsuccessful programs must be "debugged" until they satisfactorily fulfill the requirements.

A second grade teacher collects biweekly examples of representative student work in a language arts portfolio. The collected student samples are reviewed with parents during mid-year conferences. The portfolio provides parents with tangible illustrations of their child's literacy development. The teacher uses the actual products, along with a developmental scale of reading and writing for the primary grades, to discuss the student's skill strengths and point out areas needing special attention.

A college engineering professor assigns his students to work in teams to design and build a self-propelled hovering vehicle that corresponds to certain specified parameters. In addition to the model, students must individually prepare a technical report related to their designs. The models and technical reports are evaluated to determine students' understanding of, and ability to apply, principles of aerodynamics. A culminating "hovering" contest is used to determine the most effective designs.

Fifth grade art students create a landscape using tempera paints. Using a skills checklist, their art teacher assesses their paintings to determine their proficiency in using the medium. He also assesses their understanding of the use of compositional elements for creating an illusion of depth. Individual student conferences are arranged to provide feedback.

A middle school science teacher reviews her students' laboratory reports to determine their effectiveness in applying the experimental procedures and the accuracy of their data collection. Her written comments in the margins point out errors and offer specific suggestions for improvement. The reports are returned, discussed, and filed in the students' science folders for future reference.

Performance Assessments

Using performance assessments, teachers are able to observe directly the application of desired skills and knowledge. Performance assessments can be among the most authentic types of student assessments since they can replicate the kinds of actual performances occurring in the world outside of school. Performances have been used widely to assess learning in certain disciplines, such as vocal and instrumental music, physical education, speech, and theater, where performance is the natural focus of instruction. However, teachers in other subjects routinely can include performances, such as oral presentations, demonstrations, and debates, as part of a broad array of assessment methods.

Many teachers have observed that students are motivated to put forth greater effort when they peform before "real" audiences of other students, staff, parents, or expert judges.

As with product assessments, teachers apply criteria to evaluate student performances through the use of scoring tools, such as rubrics, rating scales, or checklists. The evaluation of performances becomes instructionally valuable when students apply the scoring tools for peer- and self-evaluation. Such involvement helps students to internalize the elements of quality embedded in the criteria. Many teachers have observed that students are motivated to put forth greater effort when they perform before "real" audiences of other students, staff, parents, or expert judges. In addition to the influence on students, schools often benefit from positive public relations when students perform for the community.

Despite their positive features, performance assessments can be time- and labor-intensive for students and teachers. Time must be allocated for rehearsal as well as for the actual performances. The evaluation of performances is particularly susceptible to evaluator biases, making fair, valid, and reliable assessment a challenge.

Classroom Examples

Students in the school orchestra participate in a "dress rehearsal" two weeks before the public performance. The music teacher works with the students to evaluate their performance during the rehearsal and identify areas of weakness. During the ensuing practices, the orchestra members concentrate on making improvements in these areas prior to the actual performance before a live audience.

A high school social studies teacher sets up an in-class debate as a culminating activity for a contemporary issues unit. Students work as part of a team to debate the issue of gun control. The teacher will rate students' performances in the debates on several dimensions including their understanding of the Bill of Rights, persuasiveness of their arguments, use of supporting factual information, effectiveness in countering rebuttals, and observance of rules of debating.

An elementary physical education teacher uses a skills checklist during the unit on introductory gymnastics to assess students' proficiency. Each student receives a copy of the checklist and works with a partner to try to successfully perform the identified skills. The completed checklists are used as one component of the culminating grade for the unit.

A high school speech teacher works with a home economics teacher in preparing students to make oral presentations to communicate the results of a nutrition research project. Using a rating scale, the home economics teacher evaluates the students on accuracy and completeness of their knowledge of the "food pyramid." The speech teacher uses a scoring rubric for delivery of an informative speech to evaluate the oral presentations.

Process-Focused Assessments

Process-focused assessments provide information on students' learning strategies and thinking processes. Rather than focus on tangible responses, products, and performances, this approach seeks to gain insights into the underlying cognitive processes used by students. A variety of process-focused assessments are routinely used as a natural part of teaching. For example, teachers may elicit students' thinking processes using oral questions such as, "How are these two things alike and different?," or by asking students to "think out loud" as they solve a problem or make a decision. Teachers may ask students to document their thinking over time by keeping a learning log. Finally, teachers can learn about students' thinking processes by observing students as they function in the classroom.

This "kid watching" method is especially well suited to assessing the development of attitudes or habits of mind, such as persistence.

Process-focused assessments are formative in that they provide diagnostic information to teachers and feedback to students. They also develop students' metacognition by heightening their awareness of cognitive processes and worthwhile strategies. Process-focused assessment methods are typically used over time, rather than on single occasions. Thus, they are rarely used in standardized, high stakes evaluations of students.

> ***Rather than focus on tangible responses, products, and performances, this approach seeks to gain insights into the underlying cognitive processes used by students.***

Classroom Examples

A high school literature teacher regularly poses oral questions to assess students' interpretation of a text. Their responses sometimes reveal misunderstandings that need clarification by the teacher.

A kindergarten teacher interviews each of her children in the beginning of the year. This informal assessment provides useful information about cognitive and linguistic development, social skills, and areas of personal interest.

A middle school social studies teacher carefully observes students to assess their cooperative skills as they work on a social studies project in learning groups. He also selects students to serve as process observers, giving them a checklist of observable indicators of cooperative skills. The teacher and student observers periodically provide feedback to the class on the effectiveness of their interactions in cooperative groups.

A high school mathematics teacher asks students to describe their reasoning processes by thinking out loud during the solution of "open-ended" problems. By listening to students as they articulate their thoughts, the teacher can identify fallacious reasoning and the appropriateness of strategy use so as to provide needed assistance.

Evaluation and Communication Methods

In addition to making choices about classroom assessment methods, teachers should consider options for evaluating student work and for communicating assessment results. As before, the selection of evaluation and communication methods is guided by a consideration of the key questions presented in the "Evaluation and Communication Methods" chart (Figure 3).

Figure 3

Evaluation and Communication Methods

EVALUATION METHODS

How will we evaluate student knowledge and proficiency?

Selected-Response Items:

- ❑ answer key
- ❑ scoring template
- ❑ machine scoring

Performance-Based Assessments:

- ❑ generic rubric
- ❑ task-specific guide
- ❑ rating scale
  - bi-polar
  - hierarchical
- ❑ checklist
- ❑ written/oral comments

EVALUATION ROLES

Who will be involved in evaluating student responses, products or performances?

Judgment-Based Evaluation by:

- ❑ teacher(s)/instructor(s)
- ❑ peers
- ❑ expert judges (external raters)
- ❑ student (self-evaluation)
- ❑ parents/community members
- ❑ other: _____

COMMUNICATION/ FEEDBACK METHODS

How will we communicate assessment results?

- ❑ numerical score
  - percentage scores
  - point totals
- ❑ letter grade
- ❑ developmental/proficiency scale
  - generic rubric
  - task-specific guide
  - rating scale
- ❑ narrative report (written)
- ❑ checklist
- ❑ written comments
- ❑ verbal report/conference

Evaluation Methods

How will we evaluate student knowledge and proficiency? The selection of evaluation methods should be determined largely by the assessment approach and the nature of the student responses to the assessment item or task. **Selected-response** items yield a single, correct or best answer. Most often we score selected-response items using a key to the correct or best answer choice. Sometimes we may ask students to "bubble in" their answers on an answer sheet that can be scanned by machine or hand-scored with an overlaid scoring template. Scoring of selected-response items is relatively quick, easy, and objective. **Performance-based assessments** result in a wider range of responses, products, or performances that reflect different strategies, varying degrees of quality, and different levels of proficiency. We rely on judgment-based methods to evaluate responses to these open-ended assessments. Five primary types of evaluation methods are used in conjunction with performance-based assessments: scoring rubrics, task-specific scoring guides, rating scales, checklists, and written and oral comments.

Some schools and districts have recognized the significant professional development benefits of providing opportunities for teachers to work together on scoring student products and performances and identifying anchors.

A **scoring rubric** consists of a fixed measurement scale (e.g., four or six points) and sets of criteria that describe the characteristics for each score point. We consider rubrics to be generic; that is, they provide general criteria for evaluating a student's performance in a given outcome area, such as persuasive writing. A single rubric is designed for use with many related assessment tasks. Scoring rubrics can be **holistic** (intended to provide an overall impression of the elements of quality and levels of performance in a student's work), **analytic** (designed to indicate the level of performance of a student's work on two or more separate elements of quality), or **primary trait** (global in nature like holistic rubrics but focused on a specific feature, such as language usage).

In contrast to generic rubrics, **task-specific scoring guides** are designed for use with a specific assessment activity (e.g., an open-ended question about a particular reading assignment or concept from a class discussion). While they also contain a fixed scale and descriptive criteria, task-specific guides cannot be used to evaluate responses to different performance tasks (Goldberg, 1993).

Both rubrics and task-specific guides are most effectively used for evaluation or instruction when they are accompanied by examples of responses for each score point. These examples or anchors provide

tangible illustrations of the various points on the scale. Perhaps the greatest advantage of rubrics and task-specific guides lies in their capacity to clearly communicate elements of quality to students and evaluators. The clarity provided by criteria and anchors assists raters in reliably evaluating student responses, products, or performances. The criteria also provide targets toward which teachers can teach and students can aim. When students internalize the criteria contained in rubrics or guides, they are better equipped to engage in self-evaluation and revision of their own work.

These evaluation methods require time to collect or develop rubrics and task-specific guides, to identify representative anchors, to develop proficiency in applying them reliably, and to use them for evaluating student products and performances. Nonetheless, some schools and districts have recognized the significant professional development benefits of providing opportunities for teachers to work together on scoring student products and performances and identifying anchors.

Responses to open-ended questions and tasks may also be evaluated using bi-polar and hierarchical **rating scales**. Both types of rating scales are widely used on questionnaires, and can be applied to performance-based classroom assessments. Figure 4 illustrates a **bi-polar rating scale**.

Such a scale might be used to respond to statements related to student responses, products, or performances; for example, "This argument is persuasive" or "This presentation achieves its stated purpose."

Figure 4

EXAMPLE OF A BI-POLAR RATING SCALE

strongly disagree	disagree	not sure	agree	strongly agree
–2	–1	0	1	2

Hierarchical rating scales, like the ones illustrated in Figure 5, can be used to evaluate student work along a continuum of different levels of quality or proficiency. Scoring rubrics and activity-specific scoring guides are specific types of hierarchical rating scales.

Checklists contain categories (i.e., specific features or dimensions) for evaluation and rating options for each category. The rating options may be

Figure 5

EXAMPLES OF HIERARCHICAL RATING SCALES

poor	fair	good	excellent
1	2	3	4

no essential elements evident	few essential elements evident	some essential elements evident	all essential elements evident
0	1	2	3

as simple as "yes" or "no" to indicate the presence or absence of each dimension, or presented as a narrow scale, such as "never," "rarely," or "frequently." Checklists are easy-to-use, efficient evaluation tools. They can be used while teaching a lesson or leading a discussion and are especially useful when observing students at work. Checklists may also be used as guides by students, individually or in groups, while they engage in performance activities.

While rating scales and checklists are simple to apply in the classroom, they generally do not provide the detailed, explicit criteria found in rubrics and task-specific guides. Thus, they are open to differing interpretations and greater subjectivity when used to evaluate student products and performances.

Written and oral comments can be effective in evaluating student work because they enable teachers to clearly and directly communicate with their students about elements of quality, expected standards of performance, areas of strengths, and needed improvements. These methods allow teachers to provide evaluative feedback to students on a personal level. Written and oral comments can require a great deal of teacher time and are especially demanding for secondary teachers working with one hundred or more students per day. The effectiveness of personal comments may be diminished if teachers provide only negative feedback (identifying errors or problems), make nonspecific positive comments that do not acknowledge particular aspects of student effort and work, or make comments that do not address all important elements of quality.

Roles in Classroom Assessment

Who will be involved in evaluating student responses, products, or performances? As always, this guiding question should be answered with assessment

outcomes, purposes, audiences, and methods in mind. The question also reminds us of the opportunity to involve others in the evaluation process. When students are engaged in applying criteria for self- and peer-evaluation, they begin to internalize elements of quality and performance standards in ways that can lead to improvements in the quality of their work and learning. Teachers may also involve other staff members, parents, or community experts in the evaluation of student products (e.g., science fair projects) and performances (e.g., public-speaking exhibitions).

Communication Methods

How will we communicate assessment results? A variety of methods can be used to communicate assessment results, including numerical scores, letter grades, verbal and written reports, scales and checklists. The choice of communication methods should be determined by assessment purposes and methods, evaluation methods, and especially the audience for the assessment information.

The choice of communication methods should be determined by assessment purposes and methods, evaluation methods, and especially the audience for the assessment information.

Numerical scores (e.g., percentage correct or number of points earned on a classroom quiz) and **letter grades** are widely used methods for communicating the results of classroom assessments. Both methods are efficient to use and succinct. However, numerical scores and grades, by themselves, do not explicitly communicate the elements of quality and standards of performance which they are meant to reflect. For example, saying that 70% correct is a "C" can mean different things on easy and difficult tasks and does not make clear what a student knows and can do. Likewise, when students are graded "on a curve," their knowledge or performance level is communicated in relation to other students in the class, not in terms of established criteria and standards.

Developmental and proficiency scales are generally more informative than numerical scores and grades because they contain descriptions of different degrees of quality and levels of performance. An example of a developmental scale for reading is displayed in Figure 6. Information about student learning presented in terms of developmental or proficiency levels can be especially meaningful to parents. With this recognition, some schools and districts have revised their report cards, especially for the primary grades, to incorporate features of developmental and proficiency scales.

Figure 6

DEVELOPMENTAL READING SCALE

Emergent Reader

- follows along in the text when adult reads
- aware of relationship of printed text to oral language
- uses picture cues when recalling story
- pretends to read; memorizes favorite stories

Beginning Reader

- reads word-for-word; struggles with unfamiliar material
- has limited sight vocabulary of one- and two-syllable words
- attempts to pronounce and figure out meaning or new words
- demonstrates comprehension of simple text
- occasionally monitors comprehension and self-corrects

Competent Reader

- reads familiar material comfortably
- has large sight vocabulary
- uses context clues to figure out meaning of unfamiliar words
- actively constructs meaning
- regularly monitors comprehension and self-corrects

Fluent Reader

- reads fluently with expression
- has extensive sight vocabulary
- readily determines meaning of unfamiliar words using context clues
- reads a wide variety of materials with understanding
- independently monitors comprehension; appropriately applies comprehension strategies

Checklists can also be effective for communicating assessment results since they present ratings on identified criteria or elements of quality. They are a quick and efficient method for providing direct and timely feedback to students. However, checklist developers must be careful to avoid poorly defined categories, such as creativity, that are open to diverse interpretations.

Written comments, **narrative reports**, **verbal reports**, and **conferences** can be effective communication methods because they provide opportunities to clearly and directly connect student effort and performance to elements of quality and standards of performance. They also allow teachers to provide more individualized and personal feedback than the other communication methods. Regrettably, the time-consuming nature of these methods often limits their use, especially for teachers at the secondary level because of the greater student-to-teacher ratio.

Conclusion

Assessment is an essential component of the teaching/learning process. Without effective classroom assessment, it is impossible for us to know whether our students are "hitting the target"—that is, learning what is important for them to learn. However, the significance of classroom assessment extends beyond the role of measuring learning. What we assess, how we assess and evaluate, and how we communicate results send a clear message to students about what is worth learning, how it should be learned, what elements of quality are most important, and how well we expect them to perform. By considering the key questions and principles presented here, teachers will be better equipped to develop and use classroom assessments that provide fair, valid, and reliable information that will inform teaching and promote learning.

References

Airasian, P.W. (1991). *Classroom assessment.* New York: McGraw-Hill.

Carlson, S. B. (1985). *Creative classroom testing: 10 designs for assessment and instruction.* Princeton, NJ: Educational Testing Service.

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd ed.). New York: Holt, Rinehart and Winston.

Cross, K.P., & Angelo, T.A. (1988). *Classroom assessment techniques: A handbook for faculty.* (Technical Report No. 88-A-004.0). Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Diez, M. (1993). *Essays on emerging assessment issues.* Washington, DC: American Association of Colleges for Teacher Education.

Ferrara, S., & McTighe, J. (1992). Assessment: a thoughtful process. In A. Costa, J. Bellanca, & R. Fogarty (Eds.), *If minds matter: A foreword to the future. (Vol. 2)* . Palatine, IL: Skylight Publishing.

Goldberg, G. (1993). *Scoring MSPAP: A teacher's guide.* Baltimore, MD: Maryland State Department of Education.

Haertel, E., Ferrara, S., Korpi, M., & Prescott, B. (1984). *Testing in secondary schools: Student perspectives.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R., Pickering, D., & McTighe, J. (1993). *Assessing outcomes: Performance assessment using dimensions of learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve America's schools.* New York: The Free Press.

Mitchell, R., & Neill, M. (1992). *Criteria for evaluation of student assessment systems.* Washington, DC: National Forum on Assessment.

National Education Association. *Student portfolios.* (1993). Washington, DC: NEA Professional Library.

Nitko, A. J. (1983). *Educational tests and measurement: An introduction.* New York: Harcourt Brace Jovanovich.

Osterlind, S. J. (1989). *Constructing test items.* Boston: Kluwer Academic Publishers.

Paulson, F.L., Paulson, P.R., & Meyer, C.A. (1991). What makes a portfolio a portfolio? *Educational Leadership,* 48, 5: 60-63.

Perrone, V. (Ed.) (1991). *Expanding student assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Shepard, L. (1989). Why we need better assessments. *Educational Leadership,* 46, 7: 4-9.

Smith, M.L. (1993). *The role of high-stakes testing in school reform.* Washington, DC: National Education Association.

Stiggins, R.J., & Conklin, N.F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: State University of New York Press.

Wiggins, G. P. (1989). Teaching to the (authentic) test. *Educational Leadership,* 46, 7: 41-47.

Wiggins, G. P. (1992). Standards, not standardization: Evoking quality student work. *Educational Leadership,* 4, 5: 18-25.

Appendix A: Glossary of Classroom Assessment Terms

Analytic Scoring—scoring procedure in which products or performances are evaluated for selected dimensions, with each dimension receiving a separate score. For example, a piece of writing may be evaluated on several categories, such as organization, use of details, attention to audience, and language usage/mechanics. Analytic scores may be weighted and totaled.

Anchor(s)—representative products or performances used to illustrate each point on a scoring scale. Anchors for the highest score point are sometimes referred to as exemplars.

Assessment—any systematic basis for making inferences about characteristics of people, usually based on various sources of evidence; the global process of synthesizing information about individuals in order to understand and describe them better (Brown, 1983).

Authentic—refers to assessment tasks that evoke demonstrations of knowledge and skills in ways that they are applied in the "real world." Ideally, authentic assessment tasks also engage students and reflect the best instructional activities. Thus, teaching to the task is desirable.

Criteria—guidelines, rules, or principles by which student responses, products, or performances are evaluated.

Criterion Referenced—an approach for describing a student's performance on an assessment according to established criteria.

Evaluation—judgment regarding the quality, value, or worth of a response, product, or performance based upon established criteria.

Formative Assessment—ongoing, diagnostic assessment providing information (feedback) to guide instruction and improve student performance.

Generalizability—the extent to which the performances sampled by a set of assessment activities are representative of the broader domain being assessed.

Holistic Scoring—a scoring procedure yielding a single score based upon an overall impression of a product or performance.

Indicator—a specific description of an outcome in terms of observable and assessable behaviors. An indicator specifies what a person who possesses the qualities articulated in an outcome knows or can do. Several

indicators are generally needed to adequately describe each outcome.

Interdisciplinary or Integrated Assessment—refers to tasks that assess students' abilities to apply concepts, principles, skills, and processes from two or more subject disciplines to a central question, theme, issue, or problem.

Norm Referenced—an approach for describing a student's performance on an assessment by comparison to a normed group.

Outcome—a goal statement specifying desired knowledge, skills/processes, and attitudes to be developed as a result of educational experiences.

Performance-based Assessment—an assessment activity that requires students to construct a response, create a product, or perform a demonstration. Since performance-based assessments generally do not yield a single correct answer or require a particular solution method, evaluations of student products or performances are based on judgments guided by criteria.

Performance Standard—an established level of achievement, quality, or proficiency. Performance standards set expectations about how well students should perform.

Performance Task—an assessment activity, or set of activities, related to one or more learning outcomes, that elicits one or more responses to a question or problem.

Portfolio—a purposeful, integrated collection of student work showing effort, progress, or achievement in one or more areas (adapted from Paulson, Paulson, and Meyer, 1991). Since they feature works selected over time, portfolios are well suited to assess student growth and development.

Primary Trait(s) Scoring—a scoring procedure in which products or performances are evaluated by limiting attention to a single criterion. These criteria are based upon the trait determined to be essential for a successful performance on a given task. For example, a note to a principal urging a change in a school rule might have persuasiveness as the primary trait. Scorers would attend only to that trait.

Proficiency—having or demonstrating a high degree of knowledge or skill in a particular area.

Reliability—the degree to which an assessment yields dependable and consistent results.

Rubric—a generic scoring tool used to evaluate a student's performance in a given outcome area. Rubrics consist of a fixed measurement scale (e.g., 4-point) and a list of criteria that describe the characteristics of products or performances for each score point. Rubrics are frequently

accompanied by examples (anchors) of student products or performances to illustrate each of the points on the scale.

Standardized—a set of consistent procedures for constructing, administering, and scoring an assessment. The goal of standardization is to ensure that all students are assessed under uniform conditions so that interpretation of their performance is comparable and not influenced by differing conditions (Brown, 1983).

Summative Assessment—culminating assessment for a unit, grade level, or course of study providing a status report on mastery or degree of proficiency according to identified learning outcomes.

Test—a set of questions or situations designed to elicit responses that permit an inference about what a student knows or can do. Tests generally utilize a paper-and-pencil format, occur within established time limits, restrict access to resources (e.g., reference materials), and yield a limited range of acceptable responses.

Validity—refers to whether or not an assessment measures what it is intended to measure.